Structure of the influenza virus haemagglutinin complexed with its receptor, sialic acid

W. Weis*#, J. H. Brown*, S. Cusack*#, J. C. Paulson†, J. J. Skehel‡ & D. C. Wiley*§

* Department of Biochemistry and Molecular Biology and Howard Hughes Medical Institute, Harvard University, Cambridge, Massachusetts 02138, USA
† Department of Biological Chemistry, UCLA School of Medicine, Los Angeles, California 90024, USA
‡ National Institute for Medical Research, Mill Hill, London NW7 1AA, UK

The three-dimensional structures of influenza virus haemagglutinins complexed with cell receptor analogues show sialic acids bound to a pocket of conserved amino acids surrounded by antibody-binding sites. Sialic acid fills the conserved pocket, demonstrating that it is the influenza virus receptor. The proximity of the antibody-binding sites suggests that antibodies neutralize virus infectivity by preventing virus-to-cell binding. The structures suggest approaches to the design of anti-viral drugs that could block attachment of viruses to cells.

THE haemagglutinin glycoprotein (HA) of the influenza virus membrane is responsible for binding the virus to cell-surface receptors during infection. The only known components of the receptors necessary for this interaction are sialic acids (Fig. 1); the virus does not bind to or infect neuraminidase-treated cells[1,2], and binding is restored by re-sialation[3] or addition of sialated glycolipid[4,5]. The region of the haemagglutinin involved in receptor binding has been deduced from studies of mutant haemagglutinins with different binding specificities to involve a pocket of amino acids at its membrane distal surface[6] (Fig. 2*a*). The pocket is surrounded by residues that vary with changes in antigenicity and is in fact the only suface exposed on the top of the HA that has not changed during the antigenic drift that has occurred in the HAs of Hong Kong influenza viruses since the pandemic of 1968 (Fig. 2*b*).

Using X-ray crystallography, we have determined the structure of the HA of a receptor binding mutant[6] for comparison with the structure of the wild-type molecule[37], and we have also determined the structures of both mutant and wild-type HAs complexed with sialyllactose, a trisaccharide receptor analogue. The results obtained provide insight into how influenza virus attaches to a cell, they allow assessment of mechanisms of infectivity neutralization by antibodies, and they contain information which will assist development of anti-viral drugs designed to block viral entry into cells.

The receptor binding site

Topographically, the binding site is a depression, the bottom of which is formed by the phenolic hydroxyl of Tyr 98 and the aromatic ring of Trp 153 (Figs 3 and 2*c*). Glu 190 and Leu 194 project down from a short α-helix to define the rear of the site with His 183 and Thr 155 (Fig. 3). Residues 134 to 138 form the 'right' side of the site, and residues 224 to 228 form the 'left' side.

Interactions among the conserved residues forming the surface of the pocket, and between these and a conserved 'second shell' of residues, appear to orient several of the surface atoms for binding to ligand (Fig. 3): a chain of hydrogen bonds links Trp 153, Tyr 195, His 183, Tyr 98, and via a water molecule, Glu 190; Trp 153 is part of an edge-to-face stack of aromatic rings[7] with Phe 147 and Phe (or Tyr) 148. This local stabilization may be important for maintaining the geometry of the site despite the numerous amino-acid substitutions observed to occur around its perimeter in the haemagglutinins of new epidemic strains which arise every few years. Patterns of local stabilization in other saccharide-binding proteins have been discussed by Quiocho[8].

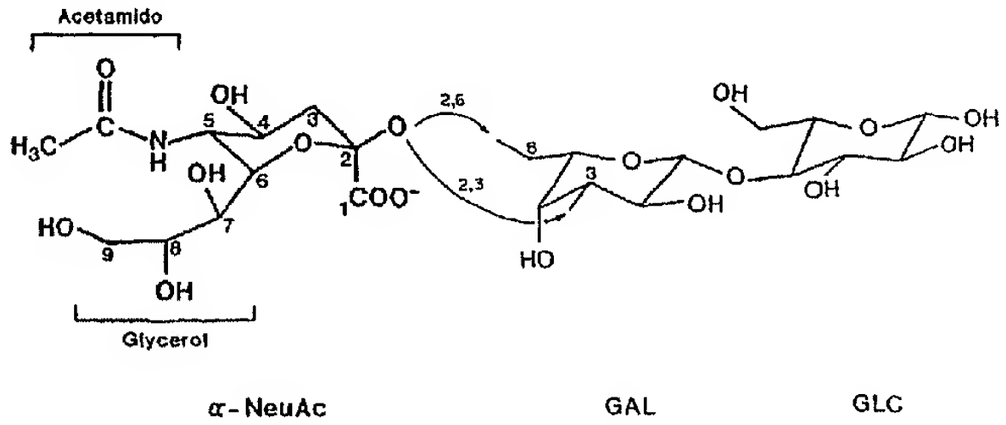

Fig. 1 Sialyllactose: α-*N*-acetyl neuraminic acid (α-NeuAc), a sialic acid, linked $\alpha(2, 6)$ or $\alpha(2,3)$ to lactose (galactose-$\beta(1,4)$-glucose), binds to the influenza HA. The α-anomers of sialic acids have their carboxylate axial to the pyranose ring at the 2 position. The acetamido and glycerol moieties of α-NeuAc are equatorial to the ring at the 5 and 6 positions, respectively.

HA-receptor analogue complexes

Complexes of wild-type HAs with $\alpha(2,6)$ linkage receptor specificity (Table 1 caption) were prepared by soaking HA crystals in $\alpha(2, 6)$ sialyllactose (NeuAc$\alpha(2,6)$Gal$\beta(1,4)$Glc), a trisaccharide with the same composition and linkages as the termini of many cell surface oligosaccharides. Table 1 summarizes details of X-ray data collection and processing, and Fig. 4*a* shows the difference electron density between $\alpha(2,6)$ sialyllactose bound to HA, and HA without sialyllactose. The density shows a flat central core with in-plane protrusions corresponding to the ring substituents of α-NeuAc, and an out-of-plane bulge at one end protruding downward towards the protein that corresponds to the axial carboxylate substituent at C2. A similar difference electron density peak is seen between the complex of $\alpha(2,3)$ sialyllactose bound to the mutant HA and the mutant HA without sialyllactose (Fig. 4*b*), indicating that α-NeuAc is bound in essentially the same way in the two sites. Three water molecules found upon crystallographic refinement of both wild-type and mutant unliganded sites (Table 1) appear to be displaced by the α-NeuAc. The fact that the water molecules are near the carboxylate and glycerol 8 and 9 positions of α-NeuAc (Fig. 4*c*) appears to explain the weak or absent density around those atoms in Fig. 4*a,b*.

No electron density corresponding to the galactose or glucose moieties of sialyllactose can be seen in either complex. This suggests that in the crystals the lactose is spatially disordered

§ To whom correspondence should be addressed.
# Present addresses: EMBL, c/o I.L.L., 156×, 38042 Grenoble Cedex, France (S.C.); Department of Molecular Biophysics and Biochemistry and Howard Hughes Medical Institute, Yale University, 260 Whitney Avenue, New Haven, Connecticut 06511, USA (W.W.).

Table 1 Crystallographic data

Variant	Compound added	d_{lim} (Å)	a, c (Å)	Number of films (source)	R_{int} whole¶	R_{int} all(>%)¶	Unique reflections (% of total)	% > 3σ	R factor#	r.m.s. deviations: 1–2 distances (Å) peptide torsions (°)	$R_{\Delta F}$**
D1112G	α(2,6)sialyl-lactose*	3.2	162.6 177.4	32 (r.a.)‡ 16 (CHESS)	0.135	0.160 (80)	63,089 (85)	64.7	0.204	0.023 3.5	0.168
G146D[52]	—	3.0	162.6 177.4	48 (r.a.)	0.115	0.115 (50)	74,765 (83)	71.4	0.209	0.023 3.3	0.169
L226Q	—	2.9	162.8 177.3	62 (r.a.) 6 (CHESS)	0.121	0.126 (70)	83,182 (83)	67.9	0.227	0.025 3.6	0.180
L226Q	α(2,3)sialyl-lactose†	2.9	162.8 176.8	32 (SSRL)‖ 16 (CHESS)	0.114	0.125 (70)	76,438 (76)	70.8	0.222	0.024 3.4	

The bromelain-released haemagglutinins were prepared as previously described[37]; all crystallize isomorphously in space group $P4_1$ (ref. 37). Crystals of D1112G were grown by microdialysis of a 35 mg ml^{-1} protein solution against 1.30–1.34 M Na_3 citrate/0.1% NaN_3/*p*H 7.5. Crystals of L226Q were grown under similar conditions, except that citrate concentration was in the range 1.34–1.38 M. For the sialyllactose studies, crystals were washed by four serial transfers through 1.35 M Na_3 citrate/0.1% NaN_3/*p*H 7.5 (D1112G) or 1.48 M Na_3 citrate/0.1% NaN_3/*p*H 7.5 (L226Q) buffer to remove contaminating viral neuraminidase, then soaked in the same buffer containing 25 mM sialyllactose for 12 to 24 h before data collection. After data collection, the soaking solutions were assayed for free sialic acid[38] to ensure that residual neuraminidase had not broken down the substrate; no significant breakdown of sialyllactose occurred. Data were collected on 1° oscillation photographs, using both rotating anode (r.a.) and synchrotron X-ray sources. Films were scanned using the program SCAN12 (ref. 39), modified for the VAX by R. C. Ladner, and processed with programs provided by Dr. P. Evans of the MRC. Polarization corrections for the synchrotron films were done essentially as in Kahn *et al.*[40], using estimates of the fraction of the primary beam polarized in the horizontal direction (0.80 at SSRL, 0.82 at CHESS). Partially recorded reflections were corrected to their fully recorded equivalents using P. Evans' version of the postrefinement procedure[41]. Data sets were scaled to one another using a linear and an isotropic temperature (exponential) factor. All statistics given are for the range between 12.0 Å and the resolution limit of the data. Crystallographic least-squares refinement of the structures was performed with the program of Hendrickson and Konnert[42]. Haemagglutinins are named relative to the A/Aichi/68 (H3N2) HA, using the single letter amino acid code. L226Q, for example, is a single amino acid mutant with 226 changed from L to Q: residues are numbered 1–328 in HA_1 and 1,001 to 1,175 in HA_2. D1112G and G146D both have wild-type receptor binding sites; their substitutions are remote from the site. They were studied in place of the wild type for practical reasons. The wild-type virus preferentially binds erythrocytes derivatized with α(2,6) linked sialic acid, while the L226Q mutant preferentially binds α(2,3) linked sialic acids[2].

* Isolated from human milk (Sigma; 85% α(2,6) isomer, 15% α(2,3).

† Isolated from bovine colostrum (Sigma; 85% α(2,3) isomer, 15% α(2,6)).

‡ Rotating anode Elliot GX-6 at 39 kV×19 mA or GX-13 at 39 kV×39 mA, CuKα radiation, 100 μm focus cup, Franks double mirror optics. Exposure time, 12 h per 1° film. Temperature, 4 °C.

§ Cornell High Energy Synchrotron Source. Wavelength, 1.557 Å (from postrefinement). Bent crystal monochromator (Si[111]) horizontal focus, mirror vertical. Exposure time, 40–80 s per 1°. Temperature, 20 °C.

‖ Stanford Synchrotron Radiation Laboratory. Wavelength, 1.5418 Å. Optics as CHESS. Exposure time, 3–6 min per 1°. Temperature, 20 °C.

¶ $R_{int} = \sum_{hkl}\sum_{obs}|I^{hkl}_{obs} - \langle I^{hkl}\rangle / \sum_{hkl}\sum_{obs} I^{hkl}_{obs}$. The $R_{int,whole}$ column gives the R factor on fully recorded intensities before postrefinement, and the $R_{int,all}$ column gives the R factor on intensities after postrefinement, including all reflections with percent recorded greater than the value given in parentheses.

\# $R = \sum_{hkl}||F_{obs}| - |F_{calc}||/\sum_{hkl}|F_{obs}|$.

** $R_{\Delta F} = \sum_{hkl}||F^{hkl}_1| - |F^{hkl}_2||/\sum_{hkl}|F^{hkl}_1|$, where 1 and 2 refer to the data sets on the lines immediatedly above and below the entry in this column.

when bound, an observation consistent with binding studies done in solution (see below). After crystallographic least-squares refinement, only the C6′ and C5′ atoms linking α-NeuAc to Gal could be placed in the α(2,6) sialyllactose complex (that is, in $2F_o - F_c$ maps).

Neither difference Fourier map shows any evidence of conformational changes in the protein upon ligand binding.

Atomic model

We present an atomic model for how the HAs of a wild-type (Fig. 4*c*) and a mutant (Fig. 2*c*) influenza virus bind sialic acid, based on interatomic distances (Fig. 5) obtained from restrained least-squares refinement of the four structures listed in Table 1. The features of the model are found in both the wild-type and mutant sialyllactose complexes, but at the 3Å resolution limit imposed by the HA crystals, it is not possible to define with complete certainty all hydrogen bonds or other intermolecular interactions. A discussion of supporting evidence for the model is given after a brief description.

The essential features of the model are that sialic acid is bound in the site with one face of the pyranose ring towards the bottom of the depression, and with the other face exposed to solution. Each of the ring substituents unique to α-NeuAc (Fig. 1) interacts with the protein, while the ring atoms and the 4-hydroxyl do not. One carboxylate oxygen, the acetamido nitrogen, and the 8- and 9-hydroxyls of the glycerol face into the site and hydrogen bond with conserved side chain and main chain polar atoms (Fig. 5*a*, *b*); the acetamido methyl group is centred over and in van der Waals contact with the six-membered ring of conserved Trp 153; Trp 153 and conserved Gly 134, Leu 194 and His 183 form a non-polar surface complementary to and in contact with a non-polar surface on sialic acid formed by C9, C7 and the acetamido methyl group. The hydroxyls at positions 4 and 7 (Fig. 1) and the acetamido carbonyl oxygen face toward solution; the 7-hydroxyl and acetamido carbonyl may form an intramolecular hydrogen bond as observed previously in crystals of the β-NeuAc methyl ester[9].

About 66% (310 Å^2) of the solvent-accessible surface[10] of the sialic acid is buried upon binding to the HA, and all of the atoms making hydrogen bonds to the protein (the 8 and 9 hydroxyls, one carboxylate oxygen, and the acetamido nitrogen) are completely inaccessible to solvent when bound. This suggests that these atoms form stronger hydrogen bonds to the protein than to water, as regions isolated from bulk solvent tend to have smaller dielectric constants[11]. The interaction of the carboxylate with a main chain amide proton (N137) rather than with a positively charged residue appears to be another example of the 'solvation' of buried charges by peptide bonds[12], although the partial solvent accessibility of the other carboxylate oxygen may diminish this effect by dispersing the charge.

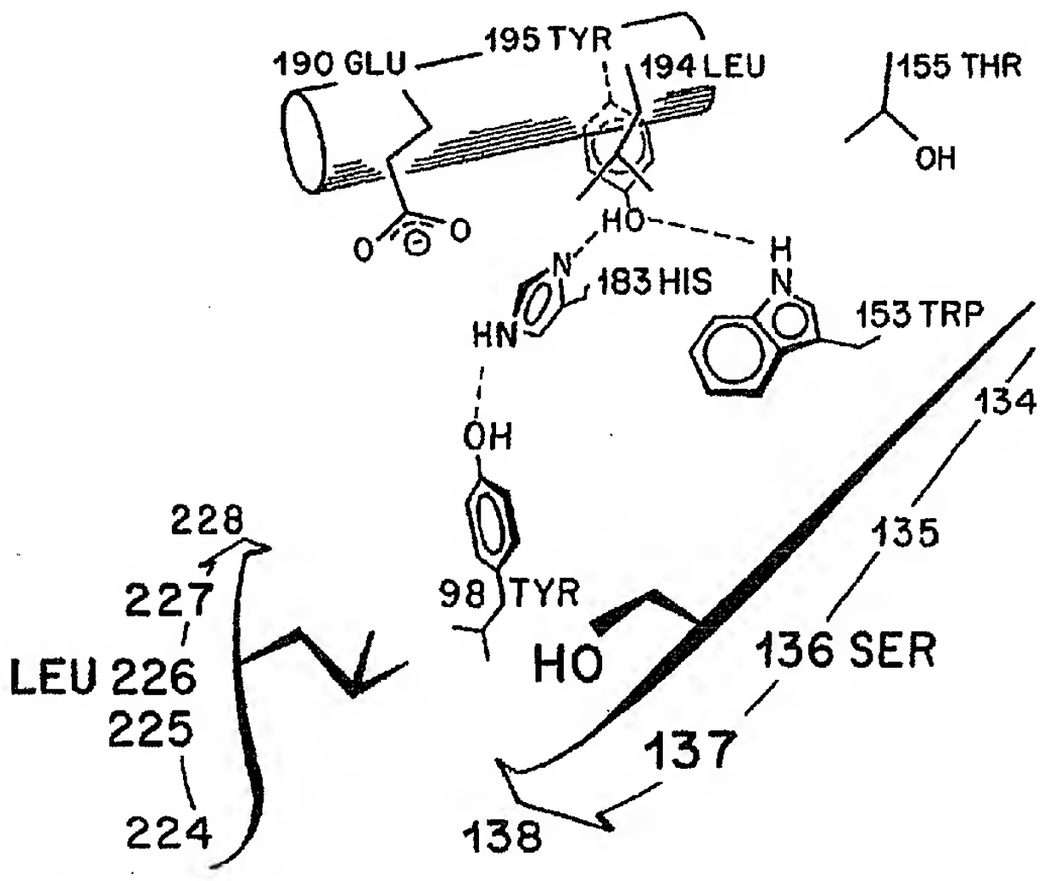

Fig. 3 The receptor binding site showing hydrogen bonding among conserved residues in site. See text for details.

Some of the observed interactions are similar to those in the recognition of *N*-acetyl groups by wheat germ agglutinin (WGA)[13,14] and lysozyme[15]: in WGA the methyl group contacts the center of a tyrosine ring, and in lysozyme, an amide proton hydrogen bonds to a main chain carbonyl oxygen.

Mutant binding site

The amino acid substitution of glutamine for leucine at 226 (L226Q) reduces the affinity of the HA for sialosides with $\alpha(2,6)$ linkages relative to $\alpha(2,3)$ linkages[6,16,17]. Recent proton nuclear magnetic resonance (NMR) measurements of the binding of $\alpha(2,6)$ and $\alpha(2,3)$ sialyllactose to the mutant HA (N. Sauter, M. Bednarski, B. Wurzburg, G. Whitesides, J. Skehel and D. Wiley, unpublished) indicate an affinity reduction of about two-fold (or only about 0.4 Kcal mol^{-1} difference in binding energy). No direct X-ray data are currently available to help explain this reduction in affinity. In the wild-type plus $\alpha(2,6)$ sialyllactose and mutant plus $\alpha(2,3)$ sialyllactose complexes (which have approximately equal affinities in solution), however, none of the linkage atoms (O2 of NeuAc, C6′, C3′ of Gal) are near the side chain at position 226. Thus the affinity reduction does not appear to be determined by a direct interaction between the substituted side chain and the linkage atoms. Instead, the affinity difference is more likely to result from

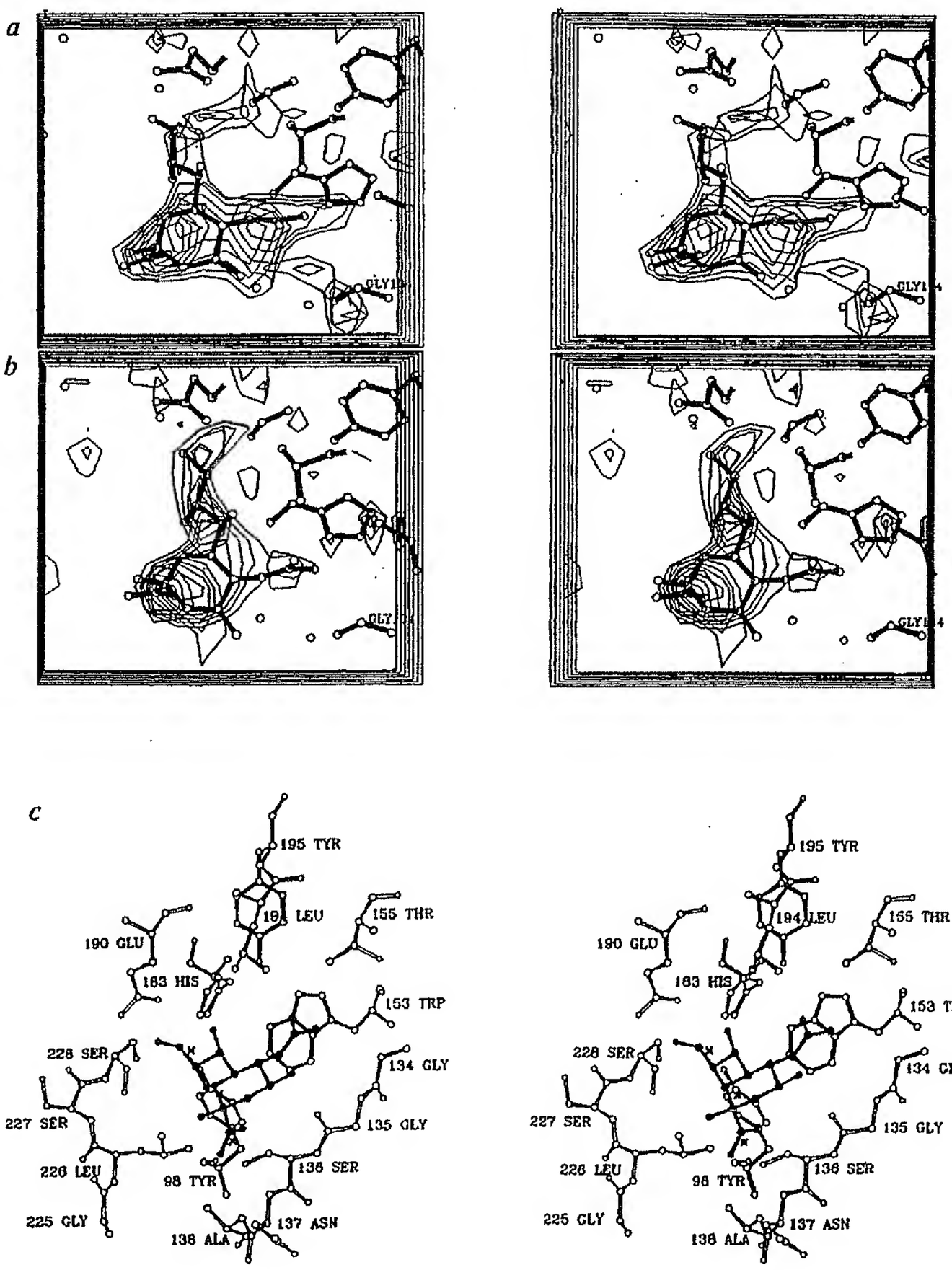

Fig. 4 Stereo pairs of α-NeuAc fitted to difference electron-density maps of wild-type (*a*) and L226Q mutant (*b*) site. See text for interpretation. The α-NeuAc model shown is based on the crystal structure of the β-anomer[47] with the carboxylate moved to the axial position. Maps have been averaged about the molecular three-fold symmetry axis, using the programs of Bricogne[48]. *a*, Difference density between wild-type site (D1112G) plus $\alpha(2,6)$ sialyllactose and the uncomplexed site. The peak shown is the only significant peak on the map aside from a peak at 1112 due to known sequence differences[49]. (The map is calculated with coefficients $(F_{obs}^{D1112G+(2,6)sialyllactose} - F_{obs}^{G146D})$ and phases from a refined G146D model[50]; contoured at 1.75 to 3.25σ above the mean density, in steps of 0.5σ.) *b*, Difference electron density between mutant L226Q binding site plus $\alpha(2,3)$ sialyllactose and the uncomplexed mutant site. The peak shown is one of two significant peaks on the map. The other is a spherical peak, smaller than a monosaccharide, located in an interior cavity on the molecular three-fold symmetry axis[49]. (The map is calculated with coefficients $(F_{obs}^{L226Q+(2,3)\ sialyllactose} - F_{obs}^{L226Q})$. The phases are calculated from a model of L226Q based on the G146D model[50], refined for 13 cycles of restrained least squares refinement[42] against the L226Q+sialyllactose data (Table 1). The map is contoured at 2.0 to 3.5σ above the mean density, in steps of 0.5σ.) *c*, Wild-type receptor site with sialic acid, stereo pair. Residues with solid bonds are conserved in all known HA sequences from natural isolates. Sialic acid is shown with solid spheres. Three water molecules displaced by sialic acid binding are indicated by X's. Criteria for water molecule positioning: (1) present in three-fold averaged and unaveraged $F_0 - F_c$ maps: (2) present in both wild type (G146D) and mutant (L226Q) $F_0 - F_c$ maps; (3) positions and temperature factors refined stably in both crystals; (4) contour level above any significant peaks in bulk solvent region; (5) reasonable distances to hydrogen bond donor and acceptors. Diagram drawn with ARPLOT[51].

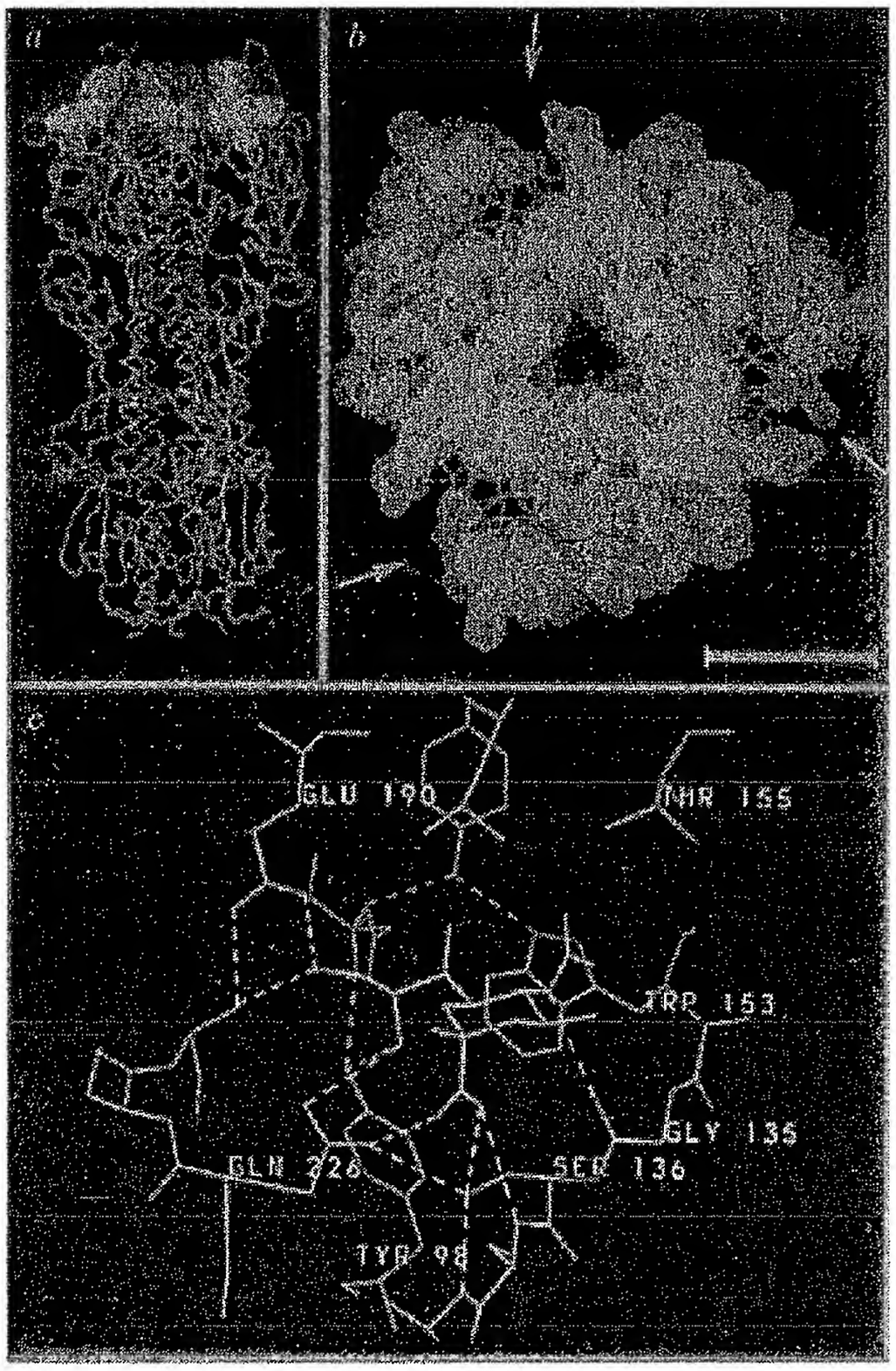

Fig. 2 The influenza virus haemagglutin binding sites. *a*, Location of the three receptor binding sites on the influenza haemagglutinin. The HA is a trimer composed of three HA_1 chains (blue) and three HA_2 chains (red) (only α carbons are shown). Each site is composed of conserved residues within a single HA_1 monomer; these residues are shown as green dot surfaces. *b*, Top view of the haemagglutinin trimer illustrating antigenic variation around the three receptor-binding sites (white arrows). Only the α carbons (coloured as in *b*) are shown. Residues in yellow are those of the A/Aichi/68 HA that have varied between 1968 and 1982 in natural isolates, and also those from antigenically distinct viruses with single amino acid substitutions selected by growth in neutralizing monoclonal antibodies. The *N*-linked carbohydrate at Asn 165, which protects the protein surface from antibody pressure[43], is also shown in yellow. The bar represents 25 Å, the approximate size of an antibody 'footprint'[25,26,27]. Antigenic sites surrounding receptor binding sites have been noted for influenza neuraminidase[44], influenza HA[45] and rhinovirus[46]. *c*, Binding site of L226Q HA with α-NeuAc bound. Yellow, carbon; blue, nitrogen; red oxygen; dashed green lines, one possible set of hydrogen bonds (see text and Fig. 5*b*). Picture generated on Evans and Sutherland PS300 with HYDRA (R. Hubbard, unpublished).

conformational differences between the mutant and wild-type proteins.

Figure 6 shows the difference in conformation between the mutant L226Q and wild-type HAs in the binding site region, based on the analysis of a difference electron density map and refinement of both structures (Table 1). Positive and negative difference electron density peaks indicate that the carbonyl oxygen and amino nitrogen of the glutamine side chain project up toward solution above one leucine methyl position, whereas the position occupied by the 'lower' methyl of the leucine is vacated (Fig. 6). A series of other positive and negative difference electron density peaks indicate small adjustments (<1 Å) of residues in the site, presumably to fill the void created by the loss of one methyl group of Leu 226 and to accommodate the new potential for hydrogen bonding of the glutamine. Tyr 98 occupies some of the space vacated by this methyl group, and Trp 153, Phe 147 and His 183 adjust to accommodate this change. The loss of a contact between Leu 226 and Ala 138 (Fig. 5*a*, *b*) "allows" the site to 'close' slightly and a new hydrogen bond to be made across the site from Gln 226 to Ser 136. Smaller but significant shifts in residue position (not shown) appear to be transmitted through positions 185 (near His 183) and 228 to 218, 219, 222 on a β strand adjacent to 226, and to residues 246 and 165 (including its *N*-linked oligosaccharide) adjacent to this strand but across the trimer subunit boundary. The conformational adjustments seen in the mutant HA are consistent with earlier suggestions[18–20] that changing residues remote from the site (244 adjacent to 246, 185, 188, 189, 218, 231 adjacent to 184) can affect receptor affinity. In some way these small conformational differences produce a reduced affinity for $\alpha(2,6)$ linked sialosides relative to $\alpha(2,3)$ sialosides.

Surprisingly, deletion of residues 224–230, which removes the left side of the site (Fig. 3) results in a viable virus, albeit with altered binding characteristics[21]. This deletion does not remove any of the conserved residues or atoms that interact directly with $\alpha(2,6)$ sialosides.

Receptor binding in solution

The interactions of each substitutent of the sialic acid ring with the HA observed in the crystal structures are consistent with those deduced from measurements in solution. The acetamido methyl proton NMR resonace exhibits a strong upfield chemical shift upon binding to the HA, characteristic of an interaction with the face of an aromatic ring (N. Sauter, M. Bednarski, B. Wurzburg, G. Whitesides, J. Skehel and D. Wiley, unpublished). This is consistent with the observation (Fig. 5*a*, *b*) that the methyl group interacts with the center of the six-membered ring of Trp 153. This ring-current shift occurs in the binding of both $\alpha(2,6)$ and $\alpha(2,3)$ sialyllactose, as well as other sialosides, to HAs of both the wild-type and L226Q mutant viruses, confirming our observation (Fig. 4*a*, *b*) that sialic acid binds in the same orientation to both HAs. A change at amino acid 155 from Thr to Tyr has been correlated with an increase in affinity for *N*-glycolyl neuraminic acid, a sialic acid with the methyl group of the acetamido substituent replaced by a hydroxymethyl group[22,23]. This is also consistent with the location of the methyl group over Trp 153 and facing towards amino acid 155 (Figs 3, 4*c*).

The interactions observed in the crystal structures at the glycosidic linkage, the carboxylate, and the 9-hydroxyl are consistent with relative binding affinities for the HA assessed by

the ability of several sialosides to delay viral adsorption to sparsely sialated erythrocytes[17]. Inhibition by α-anomeric sialosides is relatively independent of the group attached at the glycosidic oxygen (O2) of sialic acid (ref. 17 and T. Pritchett, R. Brossmer and J. Paulson, unpublished). This is consistent with the spatial disorder of lactose inferred from the crystal structures (Fig. 4*a*, *b*), which suggests that components of α-sialosides linked at O2 do not interact strongly with the HA and therefore contribute little to inhibition of viral adsorption. Evidence that the axial carboxylate at C2 of α-sialosides interacts with the protein in solution as observed in the crystal structures (Fig. 4*c*) is provided by the observations that α-anomers (axial carboxylate) inhibit 30-fold better than β-anomers (equatorial carboxylate), and that the methyl ester of α-anomers is a 10-fold weaker inhibitor than the free acid (T. Pritchett, R. Brossmer and J. Paulson, unpublished), consistent with the loss of a charged hydrogen bond and the fairly tight fit of the carboxylate in the site. The absence of any interaction between the 4-OH and the HA is consistent with the observation that acetylation at that position does not affect inhibition[24]. Conversely, the tight packing of O9 seen in the crystal structures may explain why influenza A viruses fail to bind to erythrocytes derivatized exclusively with 9-O-acetyl neuraminic acids[22]; the addition of an acetyl group of O9 would not appear to fit in the complexes that we have observed.

Neutralization of virus infectivity

In this study, the receptor binding site of the influenza virus HA has been shown to be a conserved pocket of amino acids surrounded by antigenically variable antibody binding sites. Sialic acid occupies the entire pocket, indicating that this monosaccharide is the dominant component of the influenza virus cellular receptor, and the only component for which a binding site is conserved during the antigenic evolution of the HA. These observations provide a structural proof that the influenza virus receptor is sialic acid.

The majority of antigenic sites surround the receptor-binding site, and it appears likely that antibodies bound to these sites would block receptor binding, just as *in vitro* they prevent haemagglutination. Furthermore, because of the large binding surface of an antibody (20×30 Å 'footprint'[25-27]), it is probable that even if an antibody were directed at the conserved residues in the sialic acid binding site, it would contact a number of non-conserved residues on the edges of the site and would therefore be rendered ineffective by the antibody-selected variation of those residues. Neither the precise mechanisms involved in the neutralization of virus infectivity by antibodies nor the number of antibodies required for neutralization either *in vitro* or *in vivo* have been directly determined (reviewed in refs 28–30). It is possible, for example, that virus-antibody complexes are not taken into endosomes even if limited HA-receptor interaction is sufficient for virus-cell association, and it is also possible that antibody binding to certain regions of the HA prevents the membrane fusion activity of the HA more effectively than it does receptor binding. From the results of this study, however, indicating the proximity of receptor binding and antibody binding sites, it seems reasonable to propose that a major mechanism by which antibodies neutralize influenza virus is by an effect on virus-receptor interactions.

Receptor tropism

The HA specificity for α(2,6) and α(2,3) linked sialosides of human, avian and equine influenza isolates correlates with species of origin, suggesting that receptor specificity plays a role in the biology of the virus[16]. Avian influenza viruses replicate in the intestinal tract of their host, and this tissue tropism appears to require Gln at position 226 and Gly at position 228 of the haemagglutinin[31]. The haemagglutinin of influenza virus A/Duck/Ukraine/63, a virus which may be closely related to the progenitor of the human Hong Kong influenza viruses[32],

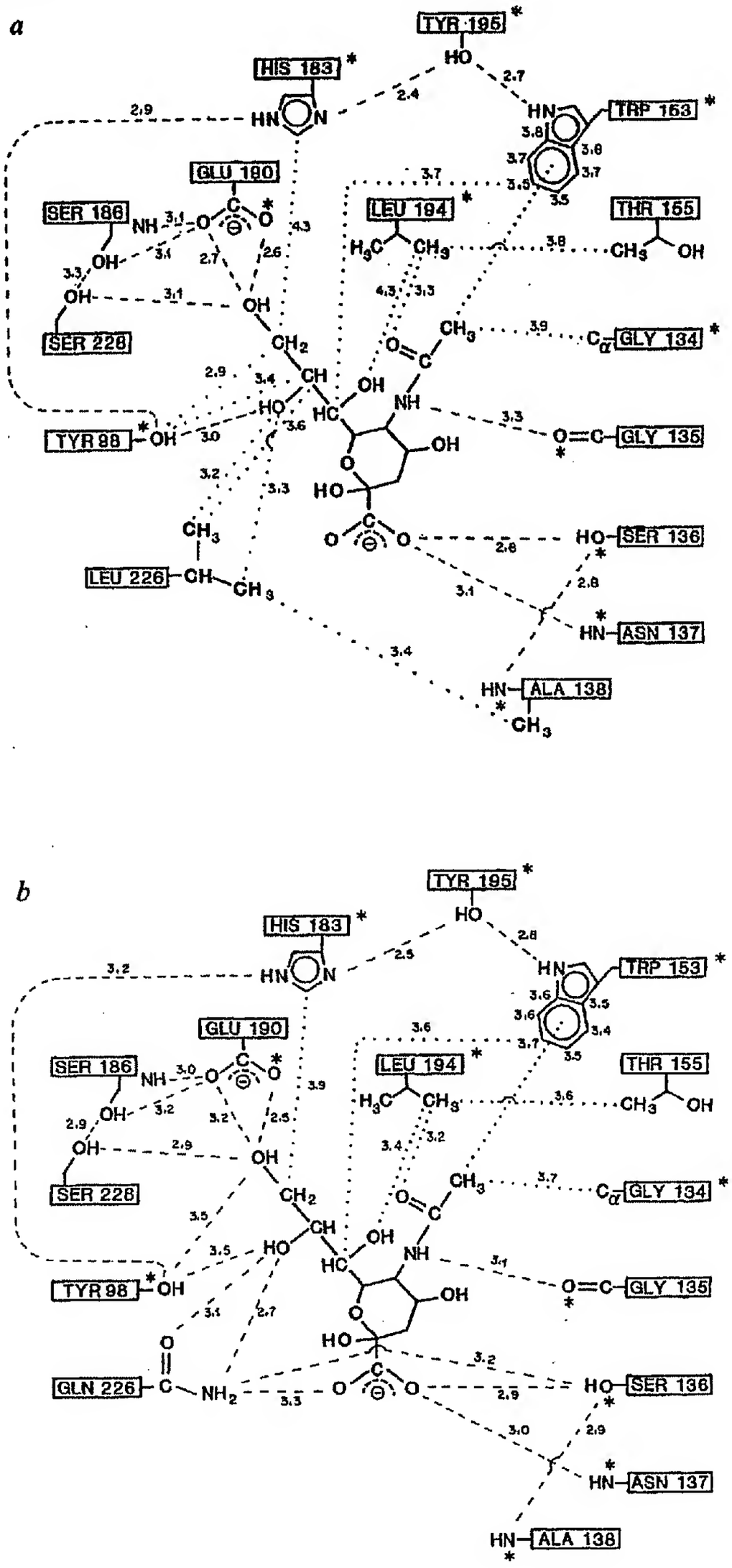

Fig. 5 Potential interactions of α-NeuAc with wild-type (*a*) and L226Q (*b*) virus HA. Potential hydrogen bonds (dashed lines) and van der Waals contacts (dotted lines) in wild type (D1112G)+α(2,6 sialyllactose) (*a*) and in L226Q+α(2,3) sialyllactose (*b*). A star (*) next to a box indicates that the residue is conserved in all known HA sequences from natural isolates; a star next to an atom means that it is main chain or conserved in substitutions. *a*, Coordinates from a refined model of D1112G+α-NeuAc. *b*, Coordinates from a model of L226Q+α-NeuAc. (The distances shown (in Å) are averages of the same distance in the three sites of the refined trimer models. Distances shown next to the atoms of the six-membered ring of Trp 153 are to the acetamido methyl group of NeuAc. Some atoms are shown with more hydrogen bonds than they can actually participate in, due to uncertainties in the refined coordinates [the coordinate error is approximately 0.3 Å].)

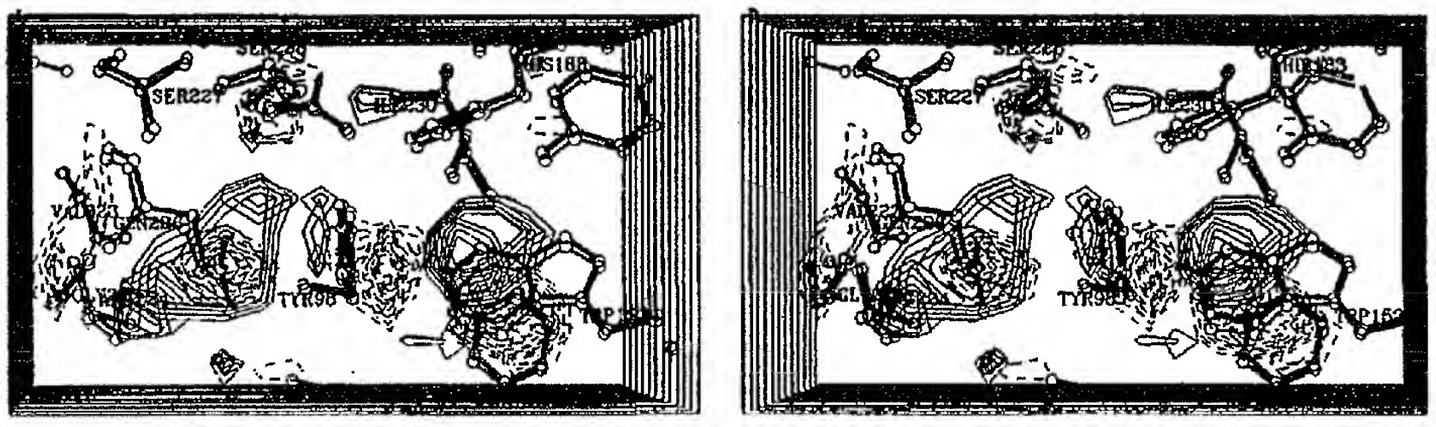

Fig. 6 Stereo pair of the difference electron density between the mutant L226Q and wild-type binding sites. The refined G146D (wild-type) and L226Q (mutant) coordinates are superimposed, with the mutant in the bolder lines. (See text for interpretation.) (The map is calculated with coefficients $(F_{obs}^{L226Q} - F_{obs}^{G146D})$ and the same phases used in the map shown in Fig. 4*a*. Map contoured at 3.0–4.5σ above the mean, in steps of 0.5σ. Positive contours are shown as solid lines, negative dashed. Averaged about the molecular three-fold symmetry axis.)

differs from the HA of the 1968 Hong Kong virus by 21 amino acids[33,34]. One of these differences is glutamine at position 226, and accordingly this avian virus, like other avian and equine influenza viruses[16,35], preferentially binds to α(2,3) linked sialoglycoconjugates[36]. A mutant of A/Duck/Ukraine/63 selected for α(2,6) binding has the single substitution Q226L[36], the reverse of the receptor site mutant studied here, indicating that even in a background of other amino acid substitutions, the conformational differences that we have observed between the single amino acid mutant and wild-type HAs influence receptor binding specificity.

Prospects for drug design

We have presented an atomic model for how influenza virus recognizes its cellular receptor, sialic acid. The details of the sialic acid–HA interaction provide a possible basis for the design of anti-viral drugs that would block viral attachment to cells. An inhibitor targeted to the conserved residues and atoms of the receptor binding site is particularly attractive, as it might be effective against influenza virus of all subtypes. This is a comparatively simple system for drug design: upon binding, no chemical reaction takes place, and apparently no conformational changes occur in the protein. Moreover, a drug targeted at the receptor binding site needs only to be presented extracellularly (to the respiratory tract), avoiding the complication of creating a compound that can cross a cell membrane. At present, however, our ability to design such drugs is limited by the resolution of our X-ray data, as well as the more fundamental problem of predicting the energetics of an intermolecular interaction from structural data.

Despite the limitations of the present structures, the interactions identified in this work lend themselves to design of HA inhibitors. Strengthening existing intermolecular hydrogen bonds by changing the electronic structure around the relevant ligand atoms may be possible, as well as adding additional atoms at favourable sites. Any attempt at drug design will have to consider the cooperativity of the virus-cell surface interaction. Even a compound with a substantially higher intrinsic affinity for the site will probably have to be presented as a polymer or on a microsurface to overcome the large effective affinity produced by the cooperative binding.

We thank Peter Rosenthal and David Stevens for excellent technical assistance, and Paul Phizackerley of SSRL and Keith Moffat and Wilfried Schildkamp of CHESS for assistance at those facilities. We acknowledge the contributions of Ian Wilson, Andy Cherenson, and Frank Escobar in our earlier studies of α-NeuAc-HA interactions. W.W. was supported by NSF and W. R. Grace predoctoral fellowships. S.C. was supported by an EMBL Visiting Fellowship, J.C.P. by the NIH and D.C.W. by grants from the NIH, NSF and synchrotron project CHESS and SSRL proposals.

1. Gottschalk, A. in *The Viruses* Vol. 3 (eds Burnet, F. M. & Stanley, W. M.) 51–61 (Academic, New York, 1959).
2. Paulson, J. C. in *The Receptors* Vol. 2 (ed. Conn, P. M.) 131–219 (Academic, Orlando, 1985).
3. Paulson, J. C., Sadler, J. E. & Hill, R. L. *J. biol. Chem.* **254**, 2120–2124 (1979).
4. Bergelson, L. D. *et al. Eur. J. Biochem.* **128**, 467–474 (1982).
5. Suzuki, Y., Matsunaga, M. & Masumoto, M. *J. biol. Chem.* **260**, 1362–1365 (1985).
6. Rogers, G. N. *et al. Nature* **304**, 76–79 (1983).
7. Burley, S. K. & Petsko, G. A. *Science* **229**, 23–28 (1985).
8. Quiocho, F. A. *A. Rev. Biochem.* **55**, 287–315 (1986).
9. O'Connell, A. M. *Acta crystallogr.* B29, 2320–2328 (1973).
10. Richards, F. M. *A. Rev. Biophys. Bioengng* **6**, 151–176 (1977).
11. Quiocho, F. A. in *Current Topics in Microbiology and Immunology* (in the press).
12. Quiocho, F. A., Sack, J. S. & Vyas, N. K. *Nature* **329**, 561–564 (1987).
13. Wright, C. S. *J. molec. Biol.* **178**, 91–104 (1984).
14. Kronis, K. A. & Carver, J. P. *Biochemistry* **24**, 826–833 (1985).
15. Kelly, J. A., Sielecki, A. R., Sykes, B. D., James, M. N. G. & Phillips, D. C. *Nature* **282**, 875–878 (1979).
16. Rogers, G. N. & Paulson, J. C. *Virology* **127**, 361–373 (1983).
17. Pritchett, T. J., Brossmer, R., Rose, U. & Paulson, J. C. *Virology*, **160**, 502–506 (1987).
18. Yewdell, J. W., Caton, A. J. & Gerhard, W. *J. Virol.* **57**, 623–628 (1986).
19. Underwood, P. A. *Arch. Virol.* **84**, 53–61 (1985).
20. Underwood, P. A., Skehel, J. J. & Wiley, D. C. *J. Virol.* **61**, 206–208 (1987).
21. Daniels, R. S. *et al. EMBO J.* **6**, 1459–1465 (1987).
22. Higa, H. H., Rogers, G. N. & Paulson, J. C. *Virology* **144**, 279–282 (1985).
23. Anders, E. M., Scalzo, A. A., Rogers, G. N. & White, D. O. *J. Virol.* **60**, 476–482 (1986).
24. Pritchett, T. J. thesis, Univ. of Calif., Los Angeles (1987).
25. Amit, A. G., Mariuzza, R. A., Phillips, S. E. V. & Poljak, R. J. *Science* **233**, 747–753 (1986).
26. Colman, P. M. *et al. Nature* **326**, 358–363 (1987).
27. Sheriff, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 8075–8079 (1987).
28. Mandel, B. *Comprehensive Virology* **15**, 37–121 (eds Fraenkel-Conrat, H. & Wagner, R. R.) (Plenum, New York, 1979).
29. Dimmock, N. J. *J. gen. Virol.* **65**, 1015–1022 (1984).
30. Dimmock, N. J. *Trends biochem. Sci.* **12**, 70–75 (1987).
31. Naeve, C. W., Hinshaw, V. S. & Webster, R. G. *J. Virol.* **51**, 567–569 (1984).
32. Laver, W. G. & Webster, R. G. *Virology* **51**, 383–391 (1973).
33. Fang, R., Minjou, W., Huylebroeck, D., Devos, R. & Fiers, W. *Cell* **25**, 315–323 (1981).
34. Ward, C. W. & Dopheide, T. A. *Biochem. J.* **195**, 337–340 (1981).
35. Daniels, R. S., Skehel, J. J. & Wiley, D. C. *J. gen. Virol.* **66**, 457–464 (1985).
36. Rogers, G. N. *et al. J. biol. Chem.* **260**, 7362–7367 (1985).
37. Wilson, I. A., Skehel, J. J. & Wiley, D. C. *Nature* **289**, 366–373 (1981).
38. Warren, L. *J. biol. Chem.* **234**, 1971–1975 (1959).
39. Crawford, J. L. Ph. D. thesis, Harvard University (1977).
40. Kahn, R. *et al. J. appl. Cryst.* **15**, 330–337 (1982).
41. Winkler, F. K., Schutt, C. E. & Harrison, S. C. *Acta crystallogr.* A35, 901–911 (1979).
42. Hendrikson, W. A. & Konnert, J. H. in *Biomolecular Structure, Function, Conformation and Evolution* (ed. Srinivasan, R.), Vol. 1, 43–57 (Pergamon, Oxford, 1981).
43. Skehel, J. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 1779–1783 (1984).
44. Colman, P. M., Varghese, J. N. & Laver, W. G. *Nature* **303**, 41–44 (1983).
45. Wiley, D. C., Wilson, I. A. & Skehel, J. J. *Nature* **289**, 373–378 (1981).
46. Rossmann, M. G. *et al. Nature* **317**, 145–153 (1985).
47. Flippen, J. L. *Acta crystallogr.* B29, 1881–1886 (1973).
48. Bricogne, G. *Acta crystallogr.* A32, 832–847 (1976).
49. Weis, W. thesis. Harvard Univ. (1987).
50. Knossow, M. *et al. Acta crystallogr.* B42, 627–632 (1986).
51. Lesk, A. M. & Hardman, K. D. *Science* **216**, 539–540 (1982).
52. Knossow, M., Daniels, R. S., Douglas, A. R., Skehel, J. J. & Wiley, D. C. *Nature* **311**, 678–680 (1984).